Causal Matrix Completion

Anish Agarwal AA5194@COLUMBIA.EDU
Columbia University
Munther Dahleh DAHLEH@MIT.EDU
MIT
Devavrat Shah DEVAVRAT@MIT.EDU
MIT
Dennis Shen DENNIS.SHEN@MARSHALL.USC.EDU
USC

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

Matrix completion is the study of recovering an underlying matrix from a sparse subset of noisy observations. Traditionally, it is assumed that the entries of the matrix are "missing completely at random" (MCAR), i.e., each entry is revealed at random, independent of everything else, with uniform probability. This is likely unrealistic due to the presence of "latent confounders", i.e., unobserved factors that determine both the entries of the underlying matrix and the missingness pattern in the observed matrix. For example, in the context of movie recommender systems—a canonical application for matrix completion—a user who vehemently dislikes horror films is unlikely to ever watch horror films. In general, these confounders yield "missing not at random" (MNAR) data, which can severely impact inference procedures that do not correct for this bias.

We develop a formal causal model for matrix completion through the language of potential outcomes and provide novel identification arguments for a variety of causal estimands of interest. We design a procedure, which we call "synthetic nearest neighbors" (SNN), to estimate these causal estimands. The SNN estimator can be seen as a combination of the synthetic controls/interventions estimator that comes from the econometrics literature with the nearest neighbor estimator that comes from the recommendation systems/matrix completion literature. The identification argument and the SNN estimator allow for (i) the probability of observing an entry of the matrix to be 0, (ii) the probability of observing an entry of the matrix to be correlated with whether other entries of the matrix are observed or not, and (iii) the missingness pattern to be correlated with the underlying outcomes of the matrix.

We prove finite-sample consistency and asymptotic normality of the SNN estimator. Our analysis also leads to new theoretical results for the matrix completion literature. In particular, we establish entry-wise, i.e., max-norm, finite-sample consistency and asymptotic normality results for matrix completion with MNAR data. As a special case, this also provides entry-wise bounds for matrix completion with MCAR data. We provide an experimental design for how to sample entries of a $m \times n$ matrix such that using the SNN procedure, we can estimate the entries of all $m \times n$ entries within approximation error δ with at most $O(\text{poly}(\delta)(m+n))$ entries.

Across simulated and real data, we demonstrate the efficacy of our proposed estimator.

Keywords: missing not at random data; causal inference; panel data; recommendation systems

Extended abstract. Full version appears as [arXiv 2109.15154]

References

A. Abadie and J. Gardeazabal. The economic costs of conflict: A case study of the basque country. *American Economic Review*, 2003.

A. Abadie, A. Diamond, and J. Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 2010.

Anish Agarwal and Rahul Singh. Causal inference with corrupted data: Measurement error, missing values, discretization, and differential privacy. *arXiv preprint arXiv:2107.02780*, 2021.

Anish Agarwal, Devavrat Shah, and Dennis Shen. On principal component regression in a high-dimensional error-in-variables setting. *arXiv preprint arXiv:2010.14449*, 2021a.

Anish Agarwal, Devavrat Shah, and Dennis Shen. Synthetic interventions. *arXiv preprint arXiv:2006.07691*, 2021b.

Anish Agarwal, Devavrat Shah, Dennis Shen, and Dogyoon Song. On robustness of principal component regression. *Journal of the American Statistical Association*, 2021c.

Gabriela Alexe, Sorin Alexe, Yves Crama, Stephan Foldes, Peter Hammer, and Bruno Simeone. Consensus algorithms for the generation of all maximal bicliques. 09 2003. doi: 10.1016/jdam.2003.09.004.

Muhammad Amjad, Devavrat Shah, and Dennis Shen. Robust synthetic control. *Journal of Machine Learning Research*, 19(22):1–51, 2018. URL `http://jmlr.org/papers/v19/17-777.html`.

Muhammad Amjad, Vishal Misra, Devavrat Shah, and Dennis Shen. Mrsc: Multi-dimensional robust synthetic control. *Proc. ACM Meas. Anal. Comput. Syst.*, 3(2), June 2019. doi: 10.1145/3341617.3326152. URL `https://doi.org/10.1145/3341617.3326152`.

Dmitry Arkhangelsky, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager. Synthetic difference in differences. Technical report, National Bureau of Economic Research, 2019.

Susan Athey, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, pages 1–41, 2021.

Jushan Bai and Serena Ng. Matrix completion, counterfactuals, and factor analysis of missing data. *arXiv preprint arXiv:1910.06677*, 2019.

Sohom Bhattacharya and Sourav Chatterjee. Matrix completion with data-dependent missingness probabilities. *arXiv preprint arXiv:2106.02290*, 2021.

Christopher M Bishop. Bayesian pca. In *Advances in neural information processing systems*, pages 382–388, 1999.

Changxiao Cai, H Vincent Poor, and Yuxin Chen. Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality. In *International Conference on Machine Learning*, pages 1271–1282. PMLR, 2020.

Changxiao Cai, Gen Li, H Vincent Poor, and Yuxin Chen. Nonconvex low-rank tensor completion from noisy data. *Operations Research*, 2021.

T Tony Cai and Wen-Xin Zhou. Matrix completion via max-norm constrained optimization. *Electronic Journal of Statistics*, 10(1):1493–1525, 2016.

Emmanuel J Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

Gary Chamberlain and Michael Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304, 1983. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1912275.

Sourav Chatterjee. Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, 43(1):177–214, 2015.

George H Chen, Devavrat Shah, et al. *Explaining the success of nearest neighbor methods in prediction*. Now Publishers, 2018.

Yuxin Chen, Jianqing Fan, Cong Ma, and Yuling Yan. Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences*, 116(46):22931–22937, 2019.

Mark A. Davenport and Justin Romberg. An overview of low-rank matrix recovery from incomplete observations. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):608–622, 2016. doi: 10.1109/JSTSP.2016.2539100.

Mark A. Davenport, Yaniv Plan, Ewout van den Berg, and Mary Wootters. 1-bit matrix completion, 2014.

Jianqing Fan, Weichen Wang, and Yiqiao Zhong. An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research*, 18(207):1–42, 2018.

Iván Fernández-Val, Hugo Freeman, and Martin Weidner. Low-rank approximations of nonseparable panel models. *arXiv preprint arXiv:2010.12439*, 2020.

Simon Funk. Netflix update: Try this at home, 2006.

Matan Gavish and David L. Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014. doi: 10.1109/TIT.2014.2323359.

David Goldberg, David Nichols, Brian M Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.

Trevor Hastie, Rahul Mazumder, Jason D. Lee, and Reza Zadeh. Matrix completion and low-rank svd via fast alternating least squares. *Journal of Machine Learning Research*, 16(104):3367–3402, 2015. URL http://jmlr.org/papers/v16/hastie15a.html.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015. doi: 10.1017/CBO9781139025751.

Nathan Kallus, Xiaojie Mao, and Madeleine Udell. Causal inference with noisy and missing covariates via matrix factorization, 2018.

Raghunandan H Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE Transactions on Information Theory*, 56(6):2980–2998, 2010a.

Raghunandan H Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from noisy entries. *Journal of Machine Learning Research*, 11(Jul):2057–2078, 2010b.

Jon Kleinberg and Mark Sandler. Using mixture models for collaborative filtering. *Journal of Computer and System Sciences*, 74(1):49–69, 2008.

Yehuda Koren. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, page 426–434, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605581934. doi: 10.1145/1401890.1401944. URL https://doi.org/10.1145/1401890.1401944.

Yehuda Koren and Robert Bell. Advances in collaborative filtering. *Recommender systems handbook*, pages 77–118, 2015.

Christina E. Lee, Yihua Li, Devavrat Shah, and Dogyoon Song. Blind regression: Nonparametric regression for latent variable models via collaborative filtering. In *Advances in Neural Information Processing Systems 29*, pages 2155–2163, 2016.

Dawen Liang, Laurent Charlin, James McInerney, and David M. Blei. Modeling user exposure in recommendation. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, page 951–961, Republic and Canton of Geneva, CHE, 2016. International World Wide Web Conferences Steering Committee. ISBN 9781450341431. doi: 10.1145/2872427.2883090. URL https://doi.org/10.1145/2872427.2883090.

Greg Linden, Brent Smith, and Jeremy York. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1):76–80, 2003.

Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

Yuping Lu, Charles Phillips, and Michael Langston. Biclique: an r package for maximal biclique enumeration in bipartite graphs. *BMC Research Notes*, 13, 12 2020. doi: 10.1186/s13104-020-04955-0.

Bingqing Lyu, Lu Qin, Xuemin Lin, Ying Zhang, Zhengping Qian, and Jingren Zhou. Maximum biclique search at billion scale. *Proc. VLDB Endow.*, 13(9):1359–1372, May 2020. ISSN 2150-8097. doi: 10.14778/3397230.3397234. URL https://doi.org/10.14778/3397230.3397234.

Wei Ma and George H Chen. Missing not at random in matrix completion: The effectiveness of estimating missingness probabilities under a low nuclear norm assumption. *arXiv preprint arXiv:1910.12774*, 2019.

Rahul Mazumder, Trevor Hastie, and Robert Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research*, 11(80):2287–2322, 2010. URL http://jmlr.org/papers/v11/mazumder10a.html.

Andriy Mnih and Russ R Salakhutdinov. Probabilistic matrix factorization. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2008. URL https://proceedings.neurips.cc/paper/2007/file/d7322ed717dedf1eb4e6e52a37ea7bcd-Paper.pdf.

Jerzy Neyman. Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes. *Master's Thesis*, 1923.

Benjamin Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12(Dec):3413–3430, 2011.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701, 1974.

Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976. ISSN 00063444.

Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. Recommendations as treatments: Debiasing learning and evaluation. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1670–1679, New York, New York, USA, 20–22 Jun 2016. PMLR. URL http://proceedings.mlr.press/v48/schnabel16.html.

Aude Sportisse, Claire Boyer, and Julie Josse. Imputation and low-rank estimation with missing not at random data. *Statistics and Computing*, 30(6):1629–1643, 2020a.

Aude Sportisse, Claire Boyer, and Julie Josses. Estimation and imputation in probabilistic principal component analysis with missing not at random data. *Advances in Neural Information Processing Systems*, 33, 2020b.

Nathan Srebro and Russ R Salakhutdinov. Collaborative filtering in a non-uniform world: Learning with the weighted trace norm. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010. URL https://proceedings.neurips.cc/paper/2010/file/67d96d458abdef21792e6d8e590244e7-Paper.pdf.

Nathan Srebro, Jason Rennie, and Tommi Jaakkola. Maximum-margin matrix factorization. volume 17, 11 2004.

Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

Madeleine Udell and Alex Townsend. Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science*, 1(1):144–160, 2019. doi: 10.1137/18M1183480.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018.

Menghan Wang, Mingming Gong, Xiaolin Zheng, and Kun Zhang. Modeling dynamic missingness of implicit feedback for recommendation. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018a. URL https://proceedings.neurips.cc/paper/2018/file/8d9766a69b764fefc12f56739424d136-Paper.pdf.

Menghan Wang, Xiaolin Zheng, Yang Yang, and Kun Zhang. Collaborative filtering with social exposure: A modular approach to social recommendation, 2018b. URL https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16058.

Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. Doubly robust joint learning for recommendation on data missing not at random. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6638–6647. PMLR, 09–15 Jun 2019. URL http://proceedings.mlr.press/v97/wang19n.html.

Yixin Wang, Dawen Liang, Laurent Charlin, and David M. Blei. Causal inference for recommender systems. In *Fourteenth ACM Conference on Recommender Systems*, RecSys '20, page 426–431, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450375832. doi: 10.1145/3383313.3412225. URL https://doi.org/10.1145/3383313.3412225.

Jiaming Xu. Rates of convergence of spectral methods for graphon estimation. *arXiv preprint arXiv:1709.03183*, 2017.

Chengrun Yang, Lijun Ding, Ziyang Wu, and Madeleine Udell. Tenips: Inverse propensity sampling for tensor completion. *arXiv preprint arXiv:2101.00323*, 2021.

Yun Zhang, Charles Phillips, Gary Rogers, Erich Baker, Elissa Chesler, and Michael Langston. On finding bicliques in bipartite graphs: A novel algorithm and its application to the integration of diverse biological data types. *BMC bioinformatics*, 15:110, 04 2014. doi: 10.1186/1471-2105-15-110.

Ziwei Zhu, Tengyao Wang, and Richard J Samworth. High-dimensional principal component analysis with heterogeneous missingness. *arXiv preprint arXiv:1906.12125*, 2019.